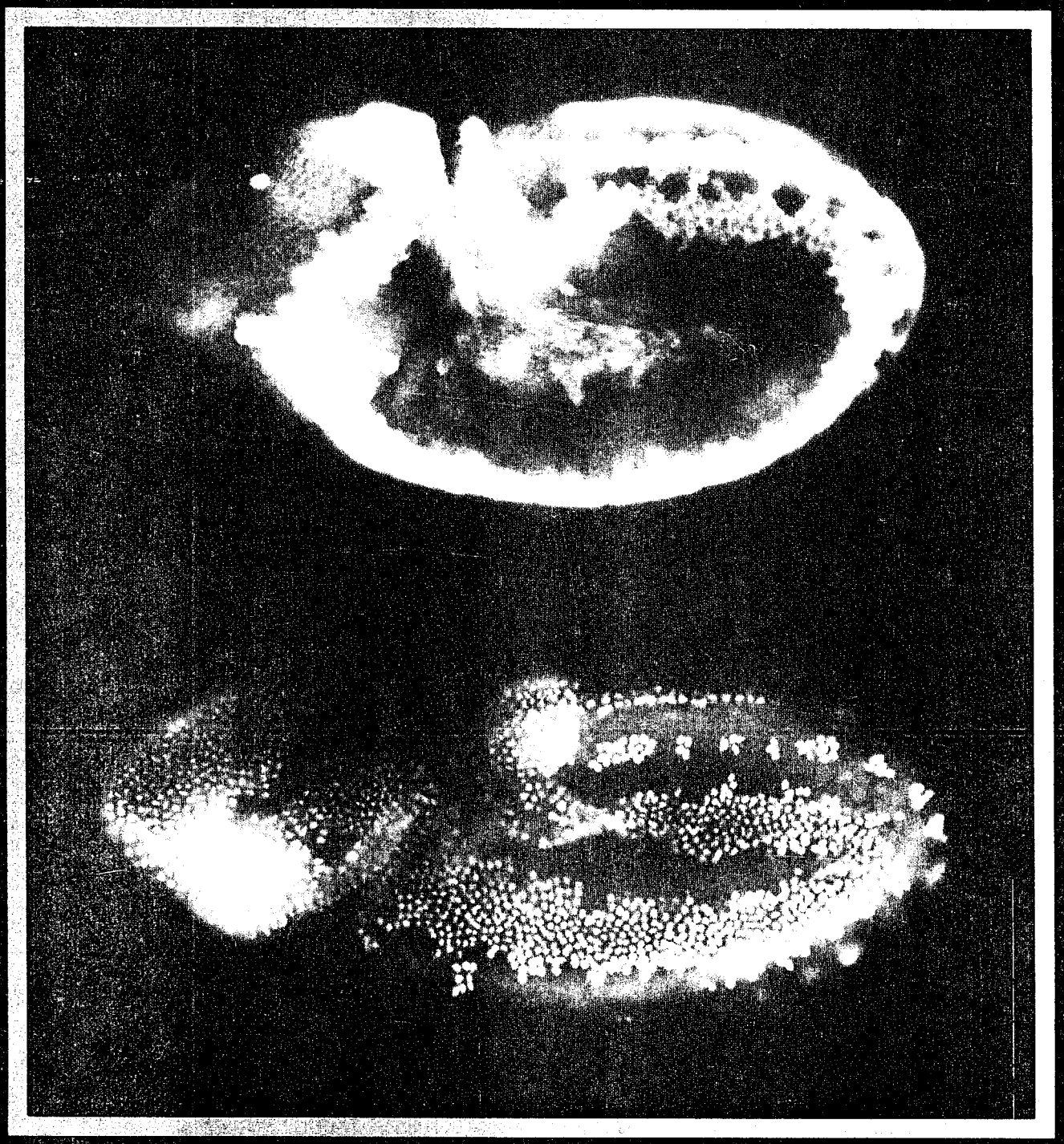
ANNUAL REPORT
1989-1990



UNIVERSITY OF CALIFORNIA, SAN FRANCISCO
SCHOOL OF MEDICINE

Department of Biochemistry and Biophysics

HAROLD E. VARMUS, M.D., PROFESSOR

MOLECULAR AND GENETIC APPROACHES TO RETROVIRUSES AND ONCOGENES

Shantanu Basu, Postdoctoral Fellow
Paul Bates, Postdoctoral Fellow
Krissy Bibbins, Graduate Student
Helene Boeuf, Postdoctoral Fellow
Delicia Caballero, Laboratory Assistant
Mario Chamorro, Staff Research Associate
Janice Clement, Secretary
Martin Hirsch, Sabbatical Visitor
Ken Kaplan, Graduate Student
Mary Jo Kelley, Administrative Assistant
Jan Kitajewski, Postdoctoral Fellow
Helen Kwan, Staff Research Associate
Andrew Leavitt, Postdoctoral Fellow
Clifford Lowell, Postdoctoral Fellow
John Mason, Postdoctoral Fellow
Tuantu Nguyen, Staff Research Associate
Suzanne Ortiz, Specialist
Neil Parkin, Postdoctoral Fellow
Vladimir Pecenka, Postdoctoral Fellow
Peter Pryciak, Graduate Student
Lily Shiue, Staff Research Associate
Supriya Shivakumar, Graduate Student
Raymond Scott, Laboratory Assistant
Peter Sorger, Postdoctoral Fellow
Karl Willert, Graduate Student
John Young, Postdoctoral Fellow

In this laboratory retroviruses serve as points of departure for studying various aspects of eukaryotic cells at the molecular level. Several properties of these viruses are particularly conducive to excursions beyond the usual confines of virology: 1) the oncogenes of retroviruses are derived from normal cellular genes (proto-oncogenes) that are often mutated in human cancers, and many proto-oncogenes are targets for retroviral insertion mutations that activate those genes in experimental tumors; 2) the DNA form of the retroviral genome (the provirus) is structurally similar to transposable elements from all types of living organisms and is inserted into chromosomes by a efficient mechanism that probably has wide currency; 3) RNA-directed DNA synthesis, a central characteristic of the retrovirus life cycle, is also fundamental to the life cycle of hepatitis B viruses, transposition by various mobile genetic elements, and the generation of certain pseudogenes, abundantly repeated sequences, and unintegrated DNA in eukaryotic and prokaryotic cells; 4) the complex pathological consequences of persistent infections by certain retroviruses, including oncogenic agents and the AIDS retrovirus, involve a wide range of interactions between viral and host macromolecules; and 5) the regulation of viral gene expression displays many features characteristic of cellular genes and some thus far unique to viruses. Consideration of these several phenomena brings us to grips with fundamental questions in eukaryotic biology.

THE RETROVIRAL LIFE CYCLE

The study of retrovirus replication has been instructive about eukaryotic cells in many ways during the past twenty years: retroviruses provided the first source of reverse transcriptase, their proviruses were the first examples of precise recombination products in eukaryotic DNA, and their strategies for gene expression include transcriptional enhancers, glucocorticoid responsiveness, alternative splicing, ribosomal frameshifting, nonsense suppression, and intricate protein-nucleic acid interactions during virus assembly. The need to understand these and other aspects of retroviruses has become more urgent with the discovery that AIDS is caused by a retrovirus, the human immunodeficiency virus (HIV).

Our current work on the retroviral life cycle is focused upon four events: the entry of virus into cells via host-encoded receptors; the integration of viral DNA into host chromosomes; the synthesis of polymerase gene products by ribosomal frameshifting; and virus assembly. Although we continue to perform most of these studies with avian and murine retroviruses, we are giving increasing attention to HIV.

Retroviral Entry: Cloning Genes for Retroviral Receptors

Entry of retroviruses into cells depends upon host-encoded transmembrane proteins that serve as receptors for viral envelope glycoproteins. The remarkable specificity of virus-host interactions has been known for over twenty years from studies of the polymorphic envelope proteins of avian retroviruses, yet little biochemical information is available about the receptors or about the nature of their interactions with viral envelope glycoproteins. Over the past few years, receptors for several animal viruses have been identified as members of the superfamily of immunoglobin genes; perhaps the most important example is the first known receptor for a retrovirus, the lymphocyte cell surface antigen CD4, a simple transmembrane glycoprotein that is required for attachment of HIV to target cells. Recently, the receptor for ecotropic murine leukemia virus (MLV) was shown to be a very different type of protein, with about fourteen transmembrane domains.

Paul Bates, John Young, and Suzanne Ortiz are attempting to clone the chicken genes that encode the receptors for multiple subgroups of avian retroviruses. In the past two years, they have accomplished an important first step by developing an assay for the receptor gene. Chicken DNA was introduced into monkey cells (presumed to lack avian virus receptors) by co-transformation with a selectable marker. A few of the co-transformed colonies were found to be infectable by avian retrovirus vectors (subgroup A) carrying another selectable marker. These cells are now being used in efforts to clone the chicken receptor gene by secondary transformation and by hybridization with subtracted cDNA probes. We ultimately expect these experiments to reveal the nature of the receptors, the basis for subgroup specificity and polymorphism, the sites of significant interaction between the receptors and envelope proteins, and perhaps the normal function of the host-encoded receptors.

The Mechanism of Proviral Integration

Like many transposable elements from plants, bacteria, yeast, and insects, retroviral proviruses can be found at many different sites in host genomes, but are always joined to host DNA at the same sites in viral DNA. The provirus contains viral genes arranged as they are in viral RNA (most commonly: *5'-gag-pol-env-3'*), flanked by long terminal repeats (LTRs) that are generated during reverse transcription and used for regulation of transcription. The LTRs terminate with short inverted repeats that form part of the att site required for integration, and the entire provirus is flanked by short direct repeats of cellular origin generated during the integration step. The only viral protein known to be required specifically for integration, the IN protein, encoded by the 3' end of the *pol* gene, was identified as an integration factor by site-directed mutants made here and elsewhere about five years ago.

Subsequent progress depended heavily upon the development of an *in vitro* system for retroviral integration, a feat achieved three years ago through a collaboration between Pat Brown, then a postdoctoral fellow in Mike Bishop's lab, and Bruce Bowerman. Bruce discovered that newly-synthesized MLV DNA was present in nucleoprotein complexes in newly-infected cells; addition of cell extracts containing those complexes to naked target DNA (e.g., lambda phage DNA) allowed efficient and correct integration of MLV DNA into the target. Although initially scored with a genetic assay, it is also possible to detect recombinants directly with a hybridization assay and thus to characterize them. We have been further assisted in the past couple of years by the production of IN proteins in yeast expression systems.

Use of these new tools has shown us that retroviral integration uses linear rather than circular DNA as substrate and that the process requires three distinct steps subsequent to completion of flush ended

linear DNA by reverse transcription. First, the IN protein recognizes and binds specifically to all sites. Second, while the viral nucleoprotein complex is still in the cytoplasm, two nucleotides are removed from the 3' ends of each strand, a step that requires viral IN protein, presumably acting as a nuclease. Then, after relocation of the nucleoprotein complex to provide access to host chromosomes, the recessed 3' ends of linear DNA are joined to newly generated 5' ends of host DNA at a site of staggered cleavage. The recombination step appears to result from a concerted cutting and joining reaction, since an external source of energy is not required for integration *in vitro*; the viral IN protein mediates this reaction as well, as recently shown by Craigie and his colleagues at NIH. Further modification of the newly linked strands may be the province of host repair systems.

The nucleoprotein complex. Bruce Bowerman has shown that the nucleoprotein complex contains the major viral capsid protein, a *gag* gene product, in addition to the expected *pol* gene products, reverse transcriptase and the IN protein, and viral DNA. At least one other *gag* protein, the matrix protein, is absent from the complex, implying that it (and perhaps other viral protein) has been removed during virus entry or reverse transcription. Bruce has also purified the complex extensively without significant loss of integration activity; thus, the complex carries with it all of the components required for integration. The DNA within the complex is susceptible to digestion with nucleases, yet the complex appears to be under structural constraints, since some antigenic determinants of the capsid protein are not recognized by anti-capsid sera.

Peter Pryciak has been investigating a prodigious source of complexes: cells infected with a simian foamy virus that accumulate as many as 10^5 copies of unintegrated viral DNA per cell. We hope to use such material for further biochemical and structural characterization of the complexes, as well as a reagent for making insertion mutations in microinjected cells of various types.

The integration protein. Shantanu Basu has expressed the MLV IN coding region in yeast and purified milligram quantities of IN protein to homogeneity. He has shown that the protein binds specifically to the att sites of MLV DNA using a "gel shift" assay with labelled synthetic oligonucleotides; the protein fails to bind to inactive mutant sites or to HIV att sites. The protein is also being tested for its ability to remove nucleotides from the 3' ends and to mediate the recombination step (using pure DNAs or mutant complexes deficient in IN protein). One of our objectives is to develop an assay that can be used to screen for anti-viral drugs. In addition, large scale purification and efforts at crystallization are underway in collaboration with Bob Stroud's group.

Andy Leavitt has initiated similar studies of the HIV IN protein, in collaboration with Phil Barr and his colleagues at Chiron. Unlike the situation with MLV, the HIV att sites have not been defined genetically and the ends of HIV DNA display very short inverted repeats. Using the gel retardation assay with HIV IN protein partially purified from yeast, Andy has shown that the recognition sites for binding are relatively large, different at each end, and probably recognized by different portions of the protein. He is currently testing for enzymatic activities of IN protein and attempting to purify the protein to homogeneity.

Host influences on integration. Peter Pryciak has been attempting to determine the mechanism responsible for restriction of an early step in MLV infection by dominant alleles at a mouse genetic locus known as *Fv-1*. Thus far, the evidence indicates that *Fv-1* restriction is not mediated by a transactive factor that can block integration *in vitro*. More likely, the restriction affects the location and amount of synthesis of linear viral DNA.

Integration into chromatin. To provide an assay that more closely simulates integration as it occurs in cells, Peter has found (with the initial help of Anita Sil, when she was a summer student here in 1989) that the MLV nucleoprotein complexes can mediate highly efficient integration into minichromosomes *in vitro*. By using well-studied minichromosomes with nucleosome-free control regions (SV40 minichromosomes from monkey cells and TRP1ARS1 minichromosomes from yeast), he is asking whether the position of nucleosomes affects integration site selection. Preliminary results with the yeast minichromosome suggest that some sites in nucleosome-containing regions are highly preferred and that favored sites occur with a periodicity of about 10 bp (i.e., one turn of the DNA helix).

Expression of Retroviral *pol* Genes by Ribosomal Frameshifting

All retroviruses synthesize reverse transcriptase (and IN protein) as *gag-pol* polyproteins that must be subsequently cleaved by a viral protease to form mature functional proteins. Yet most retroviruses encode *gag* and *pol* in different frames. When he was a graduate student here, Tyler Jacks demonstrated with *in vitro* translation assays that retroviruses solve this problem by directing ribosomes to slip back one nucleotide while decoding regions of *gag-pol* mRNAs in which the reading frames overlap. The slippage

occurs at defined frequencies (generally 5-25 percent), in response to specific sequences at the frameshift sites (e.g., A AAU UUA, or A AAA AAC), and requires secondary structural elements (e.g., a stem-loop) just downstream from the site. Although most fully explored with RSV, Tyler extended these observations with studies of the mouse mammary tumor virus (MMTV) and HIV; others have found similar frameshifting to occur during synthesis of a coronavirus RNA polymerase, the transposase of the prokaryotic element, IS1, and a subunit of *E. coli* DNA polymerase III.

Our current work on ribosomal frameshifting is addressed mainly to the role of RNA secondary structure. The strongest evidence for the requirement for a stem-loop came from disruptive and compensatory mutations of the proposed stem region of the RSV frameshift region. Mario Chamorro has obtained very similar results with the first of two frameshift regions present in the MMTV genome. However, deletion mutations made by Hiten Madhani and Tyler in RSV and MMTV genomes suggested that sequences beyond the stem-loop might also be required. Recent work in England on coronaviruses indicates that an RNA pseudoknot, rather than a simple stem-loop, forms to promote frameshifting. Mario is making mutants of MMTV and RSV to test whether pseudoknots are also formed with retroviral RNA to influence frameshifting. Neil Parkin has also begun to look for proteins that might interact with RNA secondary structure involved in frameshifting. Finally, analysis of the HIV frameshift apparatus has generated uncertainty about the role of a proposed stem-loop adjacent to the frameshift site; it appears that the frameshift site (U UUU UUA) can function in a variety of contexts, including contexts without evident potential for forming a downstream stem-loop. Mario has mutated the HIV stem-loop in its native context and found only small effects *in vitro*. Neil is now testing these mutants in cultured cells.

The many similarities between retroviruses and hepatitis B viruses prompted us to ask whether the reverse transcriptases of the latter viruses are also generated by ribosomal frameshifting. Despite the congruences in genomic layout, however, Lung-Ji Chang found that the hepatitis B viruses synthesize reverse transcriptases by internal initiation of translation on mRNA with many upstream AUGs, not by frameshifting out of the preceding core gene. Using pol-βgal fusion genes, Lung-Ji has shown that initiation is independent of the upstream core gene and not the result of a "modified scanning" mechanism, although the sequences that direct such efficient internal initiation have not been defined.

Viral Assembly

The assembly of several retroviral proteins, viral RNA, and host-derived membranes into virus particles can serve both as a model for macromolecular interactions and as a vehicle for exploring novel approaches to viral vectors. John Young, in collaboration with Paul Bates and Karl Willert (during a rotation project), has asked whether proteins containing the ectodomain of the human CD4 protein can be incorporated into the envelope of avian leukosis viruses. We expected that chimeric proteins, containing transmembrane and cytoplasmic regions derived from avian virus glycoproteins, would be favored during assembly, by virtue of interactions with the viral core proteins. But we found that normal CD4 protein was more efficiently incorporated than chimeric proteins and at least as well incorporated as the normal viral glycoproteins. We are now testing whether the CD4-bearing viruses can use the HIV envelope protein as a receptor (thereby specifically infecting HIV-infected cells), and we are looking for other host surface proteins that are efficiently assembled into retrovirus particles.

RETROVIRAL ONCOGENESIS AND PROTO-ONCOGENES

Retroviruses competent to induce tumors are conveniently grouped in two categories: those highly oncogenic agents that carry oncogenes transduced from normal cells and those less efficient agents that lack their own oncogenes. Among the first group are viruses employing over twenty distinctive oncogenes (v-*onc*'s). Each v-*onc* is derived from a cellular proto-oncogene, and these in turn are often members of gene families. A surprising number of proto-oncogenes have proved to encode growth factors, growth factor receptors, or transcription factors, but the normal functions of most proto-oncogenes are unknown, and the mechanisms by which viral versions of these genes contribute to neoplasia remain obscure. Our studies of viral oncogenes have been confined almost exclusively to v-*src*, the oncogene of Rous sarcoma virus (RSV) and the first to reveal its cellular origins.

The second group of oncogenic retroviruses, those lacking their own oncogenes, includes several viruses producing a wide spectrum of diseases. Viruses of this second type usually act as insertional mutagens, enhancing the expression of adjacent cellular proto-oncogenes as an initial step in tumorigenesis. Some of these genes have also been transduced to form retroviral oncogenes, but many were discovered as novel targets for insertion

mutation. All of our recent work in this area has been focused upon the *int*-1 gene activated by MMTV during the induction of mammary carcinomas.

The Function of the *src* Gene of RSV and its Proto-oncogene

v-*src* encodes a plasma membrane-associated phosphoprotein of 60,000 daltons (pp60$^{v\text{-}src}$) that displays protein kinase activity *in vitro* and induces phosphorylation of tyrosine residues in several proteins *in vivo*. The functionally significant target molecules have not been identified, however. The cellular progenitor, c-*src*, makes three proteins very similar to pp60$^{v\text{-}src}$ (two of which are present only in neurons as a result of differential splicing), but its physiological functions are not known. Mutations that augment the tyrosine kinase activity of pp60 appear to be required to convert c-*src* into an oncogene. Studies of c-*src* have been strongly influenced in the past few years by the recognition that it is one of about eight genes encoding very similar but differentially-regulated proteins.

Work from several laboratories is responsible for the current picture of viral and cellular *src* proteins. A short sequence at the amino terminus is a signal for addition of myristic acid, a long chain fatty acid whose presence is required for membrane localization and hence for transformation. The myristylation signal is followed by a region of about eighty amino acids that vary considerably among *src* family proteins and have no clear function. After the variable region, there are about 160 amino acids whose sequences are highly conserved among tyrosine kinases that lack transmembrane domains; these sequences, called src homology (SH)-3 and SH-2, are also present in other proteins implicated in growth control (such as phospholipase C-γ and GTPase activator protein) and are believed to modulate protein kinase activity and mediate protein-protein interactions. The carboxy-terminal half of the protein is devoted to the protein-tyrosine kinase activity and includes the ATP binding site, an auto-phosphorylation site, the probable active site for the kinase, and a site for tyrosine phosphory-lation (by an unknown kinase) that inhibits the *src* kinase activity. (This site, tyr-527, is frequently mutated in transforming alleles of *src*.)

c-src and mitosis. The c-*src* gene is expressed ubiquitously, but at especially high levels in platelets and neurons (where differential splicing can generate two isoforms), suggesting that the gene may have functions in both general growth control and specialized differentiation. Last year, David Morgan (who now has his own laboratory at UCSF) discovered that pp60$^{c\text{-}src}$ is phosphorylated during mitosis and *in vitro* by the protein kinase encoded by the vertebrate homologue of the cdc2 gene of *S. pombe*. Ken Kaplan (working jointly with David) has begun to explore the significance of this finding in two ways. First, he has made mutations in the mitotic phosphorylation sites of human pp60$^{c\text{-}src}$ and expressed the mutant proteins in rodent cells, hoping to learn whether the phosphorylations are responsible for the augmented kinase activity of pp60 in mitotic cells. Second, he has initiated studies of the localization of c-*src* protein during the cell cycle, to ask whether the protein associates with structures believed to govern the mitotic phenotype and whether locations change during the cell cycle. Past work with v-*src* protein has emphasized the association of *src* proteins with membranous components of the cell, especially plasma membranes and cytoplasmic vesicles. For example, last year Josh Kaplan (jointly supervised with Mike Bishop) showed that the aminoterminal half of *src* protein contained multiple determinants for peripheral, cytoplasmic, and perinuclear locations. Ken has now shown that a major portion of endogenous (or exogenous) pp60^{csrc} co-localizes during interphase with the microtubule organizing center (a juxtanuclear region, in the vicinity of centrosomes, from which microtubules radiate).

Mutations in c-src by homologous recombination. It is now possible to make mutations in vertebrate genes by homologous recombination between exogenous mutant DNA and the normal locus. We are attempting to exploit this new technology in two ways to study the functions of c-*src*. Clifford Lowell has made vectors designed to inactivate both copies of c-*src* in cultured cells. After several frustrated efforts to make null mutations in an embryonic carcinoma line chosen for its ability to undergo neuronal differentiation *in vitro*, Clifford has found that the same vectors will undergo homologous recombination in embryonic stem (ES) cells. He is now assessing the basis for this surprising variation in recombination frequency and proceeding to make an ES line in which both copies of c-*src* have been disrupted. To avoid the expected lethality of the double mutation, he has learned to express an exogenous copy of c-*src* from a retroviral provirus that can later be eliminated by counter-selection against another gene in the provirus. This method will allow him to test mutants of c-*src* (e.g., mitotic phosphorylation site mutants) for the ability to complement a deficiency of c-*src*.

Helene Boeuf is attempting to use homologous recombination to produce a mouse that cannot make

the neuron-specific, alternatively spliced isoforms of pp60$^{c\text{-}src}$. Targetting vectors with mutations that destroy the small exons that are normally incorporated into neuronal c-*src* mRNA are being used to make mutations in ES cells; it is intended to use those cells to make chimeric blastocysts and ultimately mice homozygous for the neuron-specific mutations.

c-src proteins. Over the past few years, Dave Morgan, Josh Kaplan, and Hisamaru Hirai have learned to make large amounts of purified products of various alleles of c-*src* and related genes using Baculovirus vectors to infect insect cells. The *src* proteins are enzymatically active and useful for a variety of biochemical studies (e.g. as targets for the mitotic kinase). In addition, we have been collaborating with Bob Stroud's laboratory on efforts to obtain crystals suitable for X-ray diffraction. Since there is no available structure for any protein-tyrosine kinase, solution to this problem is urgently required to help interpret the many interesting mutants of *src* that have recently been described (see below).

Site-directed SH3 and SH2 mutants. Hisamaru Hirai has introduced over thirty site-directed mutations into the SH3-SH2 region of a c-*src* gene made active for transformation by a mutation at tyrosine-527. The lesions include small deletions that cover all sequences in the region and base substitutions that change conserved or non-conserved amino acids. When expressed in chicken embryo fibroblasts, the mutant alleles produce a wide variety of phenotypes, but the most interesting mutations are those that enhance the oncogenic and kinase activities of pp60. (These mutations also activate pp60 without replacing tyrosine-527.) Several of the SH2 and SH3 mutants behave differently in mouse cells than in chicken cells, some being more active (as measured by cell transformation or by kinase assay) in chicken cells, and others more active in mouse cells.

Two mutants with single aminoacid changes in the most highly conserved part of SH2 have particularly dramatic phenotypes: their products show enhanced transformation and kinase activities when expressed in chicken cells, but fail to transform mouse cells and have reduced kinase activity *in vitro* when made in mouse cells. (Nevertheless, one of the mutants induces a pattern of phosphotyrosine-containing proteins in mouse cells that is similar to that seen in *src*-transformed cells.) Of special note, Hisamaru has shown that both mutants can suppress transformation by an active *src* allele in trans in a dose-dependent fashion; suppression is accompanied by degradation of the transformation-competent protein. These mutants are being exploited in several ways: to look biochemically and genetically for host factors that govern the catalytic and oncogenic activities of the proteins in the two cell types; to test candidate substrates for tyrosine phosphorylation in the two cell types; and to improve the efficiency of the trans-dominant negative effects, in hopes of interfering with the actions of normal c-*src* and other members of the gene family.

MMTV and the *int-1* Gene

The *int*-1 gene was discovered in 1982 when Roel Nusse, then a post-doctoral fellow in our group, used the technique known as "transposon tagging" to identify genes that serve as targets for insertion mutation during mammary tumor induction by MMTV. About 75 percent of tumors in C3H mice have MMTV insertion mutations that activate expression of *int*-1. (In mammary tumors in some other mouse strains, additional genes have been isolated as targets for MMTV insertion mutations; the best-studied of these, the *int*-2 gene, is a member of the fibroblast growth factor gene family.) Proviruses are found in the *int*-1 locus on either side of the coding sequence; they are generally pointed away from the transcribed region and hence act as enhancers rather than promoters of *int*-1.

A few years ago, Roel Nusse's group in Amsterdam (now at Stanford) discovered that the *Drosophila* homolog of *int*-1 is the segment polarity gene, *wingless*. Furthermore, mammals contain several genes closely related to *int*-1 and homologs have been cloned from a wide range of vertebrate species and nematodes (see below). To simplify classification and avoid confusion with unrelated genes such as *int*-2, the *wingless*-type, *int*-1-related genes have been grouped into the *wnt* gene family, with *wnt*-1 as the prototype. (The terms *int*-1 and *wnt*-1will be used interchangeably in this year's report.)

The protein product of mouse int-1. The nucleotide sequence of *int*-1 cDNA predicts a protein of 370 amino acids, with a signal peptide, a cysteine-rich carboxyterminus, and four potential N-linked glycosylation sites. Over the past few years, we have shown that the primary product is subject to multiple modifications in cultured cells, including proteolytic cleavage and several glycosylations, producing at least five distinguishable forms of *int*-1 protein in the secretory pathway. Recent studies elsewhere reveal that at least one of these proteins is secreted and associated with extracellular matrix proteins and the cell surface. Jan Kitajewski has now found that *int*-1 proteins are associated with a 78 Kd protein known as

BiP, which binds to certain secretory proteins in the endoplasmic reticulum to either assist secretion or remove incorrectly folded proteins.

To investigate the significance of the complex processing of *int*-1 proteins, John Mason has produced a collection of site-directed mutants of *int*-1 cDNA, with lesions that affect the signal peptide, remove potential cleavage and glycosylation sites, and add or subtract cysteine residues. Using an assay for transformation of cultured mammary epithelial cells, he has found that each of the glycosylation and cleavage sites are dispensible for transforming activity, but retention of the signal peptide is required. Most changes involving cysteine residues are likewise important, as implied by the extraordinary conservation of their pattern throughout the *wnt* gene family. The processing and export of these and other mutant proteins (e.g., their interactions with BiP) are now being tested.

Since *int*-1 protein appears to be a secreted glycoprotein, it is presumed to act by interacting with an as yet unidentified cell surface receptor. Jan Kitajewski is attempting to study the putative receptor through the use of purified *int*-1 proteins synthesized in baculovirus expression systems. Despite considerable difficulties posed by relatively low levels of production, poor solubility, and aggregation, Jan has prepared adequate amounts of *int*-1 protein for iodination and binding experiments and for tests of biological activity. In addition, he has obtained provisional evidence for the possibility that the *int*-1 receptor is a transmembrane protein-tyrosine kinase.

The int-1 gene as a mammary oncogene. A few years ago, when he was a fellow in our laboratory, Tony Brown showed that MLV vectors carrying the *int*-1 gene could induce changes in the growth properties and morphology of a cultured mammary epithelial cell line, but did not affect several other cell types. Because *int*-1 did not produce a fully oncogenic phenotype in the cultured mammary cells, Ann Tsukamoto (in collaboration with Rudi Grosschedl) constructed transgenic mice carrying an *int*-1 gene activated by an MMTV LTR. The mammary glands in both male and female transgenic mice exhibit marked epithelial hyperplasia, and the females have a high incidence of mammary carcinoma indistinguishable from the virus-induced disease. (In addition, some males have mammary carcinomas, and a few transgenic animals have developed salivary gland carcinomas.)

Although results with transgenic mice document the oncogenic potential of *int*-1, they also indicate that *int*-1 is not sufficient for tumorigenesis. Only a few cells appear to develop into mammary carcinomas after several months of age, despite the diffuse hyperplasia of mammary tissue; and the time course of tumor appearance is similar to that observed in virus-infected, non-transgenic animals, despite the pre-existence of an activated *int*-1 gene in each cell of the transgenic animals. We are therefore looking for cellular genes that might collaborate with *int*-1 during tumorigenesis and asking whether the viral genome might contribute in ways other than insertion mutation.

To address the first issue, Ann and Helen Kwan have crossed our *int*-1 transgenic mice with transgenic mice (provided by Phil Leder's laboratory at Harvard) that carry the *int*-2 gene under control of the MMTV LTR. Although the *int*-2 transgenic mice do not develop mammary carcinomas, the double transgenic progeny, both females and (especially) males, show a marked acceleration of tumorigenesis compared to the *int*-1 transgenics. To explore the mechanism of collaboration between these oncogenes, Vladimir Pecenka is following the onset of expression of the transgenes.

As another means of finding contributory oncogenes in the transgenic model, Greg Shackleford and Helen Kwan have infected *int-1* transgenic mice with MMTV. Mammary tumors appear at least one to two months earlier in virgin or breeding females after MMTV infection than in control animals, and most of the tumors have new MMTV proviruses in a pattern that implies clonal growth of an infected cell. Greg (now at USC) is attempting to identify proto-oncogenes that may have been activated by these MMTV proviruses.

To ask whether unidentified viral proteins might contribute to mammary oncogenesis, Greg has also made mutations in an unexplained long open reading frame in the MMTV LTR. Preliminary results in BALB/c mice suggest that the mutant virus occasionally induces tumors, but much less efficiently than does wild-type virus. These experiments are being repeated in mice lacking endogenous viruses closely related to MMTV (in collaboration with Bob Callahan's group at the NIH).

int-1 as a determinant of neural and germ cell development. A few years ago, Greg Shackleford found that, unlike proto-oncogenes that are expressed in many types of cells and considered to function as general regulators of growth, *int*-1 is expressed in only two situations in the mouse: early spermatids (a post-meiotic germ cell) in the testis and the developing neural tube, between days 8 and 15 of

embryogenesis. Several kinds of experiments have been initiated in efforts to understand the developmental consequences of *int*-1 expression. To identify more precisely the cells in which *int*-1 is expressed, John Mason is making transgenic mice that produce readily detectable proteins (e.g., β-galactosidase) under the control of *int*-1 transcriptional signals. At least two mouse lines express an *int*-1-beta-galactosidase transgene during embryogenesis. In one case, the pattern is clearly different from that observed by McMahon and his colleagues by *in situ* hybridization with *int*-1 probes and presumably represents a position effect on transcription. The other pattern is very similar (in both central nervous system and testes) but also subtly distinct from the reported *int*-1 pattern, and will require further study. Additional constructs are being tested to obtain mice with the desired pattern more regularly. Definition of the sequences required for correct expression will be important for attempts to ablate cells that express the gene or attempts to complement *int*-1 deficiencies in mice being engineered elsewhere.

The wnt gene of C. elegans. As an alternative approach to the function of the *wnt* gene family, we have been collaborating with Cynthia Kenyon's laboratory to study the nematode homolog(s) of mammalian and insect *wnt* genes. Thus far, one *wnt* gene has been cloned by Sasha Kamb, Greg Shackleford, and John Mason, starting with degenerate *wnt*-family primers for polymerase chain reactions (PCR). Sequencing of genomic and cDNA clones by Greg and Lily Shiue demonstrates that the *C. elegans wnt* protein is about 33 percent identical to other family members, with nearly complete conservation of cysteine residues. One interesting difference is in the intron-exon organization; although most *wnt* genes have the same arrangement of exons, the *C. elegans* gene has several additional introns and only one in a conserved position.

Supriya Shivakumar is using nucleic acid hybridization and mutagenic methods to study the function of the nematode gene. Northern blotting reveals three *wnt* RNA species in unstaged worms; we do not yet know how the three are generated, and we have not yet been able to localize the transcripts by *in situ* hybridization, although the method works with more abundant RNAs. Supriya has developed a mutagenic scheme that should allow detection of *wnt* deletion mutants by PCR and isolation of mutants by sib selection; this would represent the first example of a *C. elegans* mutant obtained without dependence on the phenotype.

PUBLICATIONS

Basu, S., and H.E. Varmus (1990). Moloney murine leukemia virus integration protein produced in yeast binds specifically to viral att sites. *J. Virol.* **64:** 5617-5625.

Bowerman, B., P.O. Brown, J.M. Bishop, and H.E. Varmus (1989). A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev.* **3:** 469-478.

Brown, P.O., B. Bowerman, H.E. Varmus, and J. M. Bishop (1989). Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc. Natl. Acad. Sci. USA* **86:** 2525-2529.

Chang, L-J., D. Ganem, H.E. Varmus (1990). Mechanism of translation of the hepadnaviral polymerase (P) gene. *Proc. Natl. Acad. Sci. USA* **87:** 5158-5162.

Chang, L.-J., J. Dienstag, D. Ganem,. and H.E. Varmus (1989). Detection of antibodies against HBV polymerase antigen in hepatitis B virus-infected patient. *Hepatology* **10:** 332-335.

Chang, L.-J., P. Pryciak, D. Ganem, and H.E. Varmus (1989). Biosynthesis of the reverse transcriptase of hepatitis B viruses involves de novo translational initiation not ribosomal frameshifting. *Nature* **337:** 364-368.

Chang, L.-J., R. Hirsch, D. Ganem, H.E. Varmus (1990). Effects of insertional and point mutations on the functions of the duck hepatitis B virus polymerase. *J. Virol.* **64:** 5553-5558.

De Lange, T., L. Shiue, R.M. Myers, D.R. Cox, S.L. Naylor, A.M. Killery, and H.E. Varmus (1990). Structure and variability of human chromosome ends. *Mol. Cell. Biol.* **10:** 518-527.

Hirai, H. and H.E. Varmus (1990). Site-directed mutagenesis of the SH2- and SH3- coding domains of c-src produces varied phenotypes, including oncogenic activation of $P60^{c\text{-}src}$. *Mol. Cell. Biol.* **10:** 1307-1310.

Hirsch, R.C., J.E. Lavine, L.-J. Chang, H.E. Varmus, and D. Ganem (1990). Polymerase gene products of hepatitis B viruses are required for genomic RNA packaging as well as for reverse transcription. *Nature* **344:** 552-555.

Kamb, A., M. Weir, B. Rudy, H.E. Varmus, and C. Kenyon (1989). Identification of genes from pattern formation, tyrosine kinase and potassium channel families by DNA amplification. *Proc. Natl. Acad. Sci. USA* **86:** 4372-4376.

Kaplan, J.M., H.E. Varmus, and J. M. Bishop (1990). The *src* protein contains multiple domains for specific attachment to membranes. *Mol. Cell. Biol.* **10:** 1000-1009.

Morgan, D.O., J.M. Kaplan, J. M. Bishop, and H.E. Varmus (1989). Mitosis-specific phosphorylation of $p60^{c\text{-}src}$ by $p34^{cdc2}$-associated protein kinase. *Cell* **57:** 775-786.

Varmus, H.E. (1989). Historical overview of oncogenes. In *Oncogenes and the Molecular Origins of Cancer* (R. Weinberg, ed.), 3-44. Cold Spring Harbor Press, Cold Spring Harbor.

Varmus, H.E. (1989). Lessons from the life cycle of retroviruses. In *The Harvey Lectures, Series 83*, 35-56. Alan R. Liss, Inc., New York.

Varmus, H.E. (1989). Naming the AIDS virus. In *The Meaning of AIDS: Implications for Medical Science, Clinical Practice, and Public Health Policy* (E. T. Juengst and B. A. Koenig, eds.), 3-11. Praeger Publishers.

Varmus, H.E. (1989). Regulation of HIV and HTLV gene expression. *Genes and Development* **2:** 1055-1062.

Varmus, H.E. (1990). Retroviruses and oncogenes - I. In *Les Prix Nobel,* and *Angew. Chemie. Int. Ed. Engl.* **29,** 707-822.

Varmus, H.E. (1989). Reverse transcription in bacteria. *Cell* **56:** 721-724.

Varmus, H.E. (1989). Transgenic mice and host cell mutants resistant to transformation as model systems for identifying multiple components in oncogenesis. In *Genetic Analysis of Tumour Supression* (G. Bock and J. Marsh, eds.), 30-35. John Wiley & Sons, Ltd.

Varmus, H.E. and P.O. Brown (1989). Retroviruses. In *Mobile DNA* (Howe and Berg, eds.), 53-108. ASM Publications.

Varmus, H.E., H. Hirai, D. Morgan, J. Kaplan, and J.M. Bishop (1990). Function, location and regulation of the *src* protein-tyrosine kinase. In *Genetic Basis for Carcinogenesis: Tumor Supressor Genes and Oncogenes* (A. Knudson, Jr., et al., eds), 63-70. Japan Science Press, Tokyo/Taylor & Francis, LTD., London.

Verderame, M.F., J.M. Kaplan, and H.E. Varmus (1989). A mutation in v-*src* that removes a single conserved residue in the SH-2 domain of $pp60^{v\text{-}src}$ restricts transformation in a host-dependent manner. *J. Virol.* **63:** 338-348.

IN PRESS

Morgan, D.O., J.M. Kaplan, J.M. Bishop, H.E. Varmus (1990). Production of $p60^{c\text{-}src}$ by baculovirus expression and immunoaffinity purification. In *Methods in Enzymology*, in press.

MAJOR RESEARCH SUPPORT

Source:	National Institutes of Health
Title:	Molecular analysis of retroviruses and oncogenes
Total award period:	August 1, 1985 to July 31, 1992
Source:	American Cancer Society
Title:	Research Professorship
Total award period:	1984 -
Source:	National Institutes of Health
Title:	Structural biology and targeted drug design for AIDS
Total award period:	September 1, 1987 to August 31, 1992
Source:	National Institutes of Health
Title:	Development of new approaches to inhibit growth of HIV
Total award period:	September 1, 1988 to August 31, 1993

PERSONNEL WHO LEFT THE LABORATORY

Bruce Bowerman, Postdoctoral Fellow, University of Washington
Lung-Ji Chang, Staff Scientist, National Institute of Allergy and Infectious Disease
Titia deLange, Assistant Professor, Rockefeller University
Hisamaru Hirai, Assistant Professor of Medicine, University of Tokyo
Alexander Kamb, Postdoctoral Fellow, UCSF (Stroud lab)
David Kaplan, Staff Scientist, Frederick Cancer Research Center
David Morgan, Assistant Professor, Department of Physiology, UCSF
Greg Shackleford, Assistant Professor of Pediatrics, Children's Hospital, University of Southern California
Jan Tuttleman, Motherhood State